Biomedical Optics EXPRESS

Deep learning 2D and 3D optical sectioning microscopy using cross-modality Pix2Pix cGAN image translation

HUIMIN ZHUGE,[1] BRIAN SUMMA,[2] JIHUN HAMM,[2] AND J. QUINCY BROWN[1,*]

[1] *Department of Biomedical Engineering, Tulane University, 500 Lindy Boggs Center, New Orleans, LA 70118, USA*
[2] *Department of Computer Science, Tulane University, New Orleans, LA 70118, USA*
[*] *jqbrown@tulane.edu*

Abstract: Structured illumination microscopy (SIM) reconstructs optically-sectioned images of a sample from multiple spatially-patterned wide-field images, but the traditional single non-patterned wide-field images are more inexpensively obtained since they do not require generation of specialized illumination patterns. In this work, we translated wide-field fluorescence microscopy images to optically-sectioned SIM images by a Pix2Pix conditional generative adversarial network (cGAN). Our model shows the capability of both 2D cross-modality image translation from wide-field images to optical sections, and further demonstrates potential to recover 3D optically-sectioned volumes from wide-field image stacks. The utility of the model was tested on a variety of samples including fluorescent beads and fresh human tissue samples.

1. Introduction

In traditional wide-field fluorescence microscopy, the whole specimen volume is exposed axially to illumination. This results in the generation of fluorescence not only at the focal plane of interest, but also above and below the focus plane for thick samples. This dramatically obscures specimen details and reduces contrast due to contamination of the focal plane image with out-of-focus fluorescence emission. Optical sectioning is the process by which clear images of the focal plane are preferentially recovered from a thick sample. For example, laser scanning confocal microscopy [1–3] uses a pinhole to block out-of-focus light to restrict the collected fluorescence to that emitted from the focal plane. Multiphoton microscopy [4,5] relies on the simultaneous absorption of multiple photons, which excites molecules preferentially at the focal plane, avoiding out-of-focus emission altogether. Light sheet microscopy [6,7] also relies on the restriction of fluorophore excitation to the focal plane with a thin sheet of laser light oriented orthogonally or otherwise off-axis from the optical detection axis.

Structured Illumination Microscopy (SIM) [8,9] is a wide-field optical sectioning technique, which relies on the preferential modulation of image contrast at the focus plane, and subsequent computational demodulation to isolate photons emitted from the plane of focus. SIM involves illuminating the sample with patterned or modulated light, that efficiently modulates excitation of fluorophores in the sample only within a specific axial confinement volume. By acquiring multiple wide-field images with the modulation pattern adjusted between each exposure, an optically sectioned image can be retrieved using a variety of simple algorithms [10,11]. The advantage of SIM is that it is fast and relatively simple, being itself a wide-field imaging technique. However, it still needs specialized hardware and requires multiple images to be obtained of each area of the sample, increasing the complexity and limiting speed compared to traditional wide-field fluorescence microscopy.

#439894 https://doi.org/10.1364/BOE.439894
Journal © 2021 Received 9 Aug 2021; revised 1 Nov 2021; accepted 7 Nov 2021; published 12 Nov 2021

An alternative approach to achieve optical sectioning images with low cost and high quality, is to employ innovative machine learning methods for image-to-image translation, for which there are only a few emerging studies. The goal of such multimodal unsupervised image-to-image translation methods is to learn the mapping from the input image to the corresponding output image. Ling, et al. [12] showed that a convolutional neural network (CNN) can be used to reconstruct a super-resolution SIM image with three instead of nine raw images. Jin, et al. [13] obtained a five-fold reduction in the number of raw SIM images required for super-resolution purposes using a U-Net [14,15], and they also showed the ability to reconstruct images from noisy inputs, recovering signal from poor quality inputs. However, in these cases, patterned images were still required as future inputs to the model, and none exploited the 3D optical-sectioning capability of SIM. Christensen, et al. [16] recently demonstrated an approach for reconstruction of super-resolution SIM images using simulated training data and transfer learning, however this work did not focus on the optical sectioning capability of SIM. Zhang, et al. [17] proposed a CNN architecture to reconstruct optically sectioned images using wide-field images as input and trained with paired pixel-aligned SIM images. Ning, et al. [18] further generalized the 2D CNN model to handle 3D whole-brain imaging at a single-neuron resolution. Using these approaches, optically-sectioned images can be acquired from traditional wide-field fluorescence microscopy, enabling one to bypass the implementation of additional components for the optical system. However, all of the CNN models mentioned above, are aimed at reducing the mean squared error (MSE) only between ground truth and predicted images, which could make the predicted result overly smooth, lacking high-frequency details [19].

In addition to the CNN approaches above, other image-to-image translation [20–23] approaches using cGAN have been studied over the past few years in computer vision. Isola, et al. [24] proposed a Pix2Pix cGAN model containing a U-net [14,15] generator plus a discriminator to perfectly compute the translation between spatially aligned pairs of images. Since a GAN [25] has the adversarial relationship between the generator and discriminator, they can achieve a better result in this type of multi-modal transformation task. The cGAN model has been employed on various problems in the field of microscopy, including histological stain domain transfer [26–30], super-resolution over a large field of view (FOV) [31,32], image synthesis and semantic segmentation [33,34], and image restoration (denoising, refocusing, standardization, normalization) [35–38]. Overall, cGANs can learn the complicated distributions of the data, and work well with a lack of data. To our knowledge, there have been no reports of the application of cGANs in the field of optical sectioning.

Herein, we contribute a cGAN model trained with paired wide-field and SIM images to enable both 2D and 3D optical sectioning using only wide-field images as inputs, and furthermore provide initial results that the model is generalizable to wide-field microscopes without dedicated SIM illumination hardware. Specifically, we trained a cGAN model to reconstruct optically-sectioned SIM images from uniformly-illuminated wide-field input images. We used mean square error (MSE), structural similarity index measure (SSIM), peak signal to noise ratio (PSNR), and mutual information (MI) as evaluation metrics to assess the result compared to ground truth. First, we implemented a 2D cGAN model [24] to recover optically-sectioned planes from in-focus wide-field images, and tested it on images from fluorescent bead phantoms and human tissues. As we show in the results section, the cGAN model produces better results than previously described CNNs for this type of task both in image quality and quantitative metrics. Then, we developed a 3D model, which was tested on fluorescent bead phantoms to evaluate the ability to recover true 3D images using this approach. Furthermore, additional contributions include a method to alleviate "checker board" effects caused by the nature of the deconvolution operation, and a pipeline to achieve seamless stitched images from model-predicted patches, to enable the method to be applied in large-scale mosaic imaging applications.

2. Materials and methods

2.1. Microscopy

For training the model, we employed a custom-built structured illumination microscopy system [11]. The system comprises a custom-built SIM module attached to a commercial modular automated epi-fluorescence microscope platform (RAMM, Applied Scientific Instrumentation), and incorporates a 7 mm/s motorized XY specimen stage and a motorized Z objective positioner, a fast ferroelectric liquid-crystal-on-silicon spatial light modulator (SLM) (3DM, Forth Dimension Displays), and a fast scientific complementary oxide semiconductor (sCMOS) camera (Orca Flash 4.0 v2, Hamamatsu). The illumination was provided by a fiber-coupled multi-line wide-field laser illuminator (LDI-6, 89 North). In this work, images were collected using a Nikon Plan Apo 10X 0.45NA objective lens. A schematic of the SIM system is shown in Fig. S1.

To test the generalizability of the trained model to wide-field images collected with different microscopes, we used a standard Nikon TE2000 inverted fluorescence microscope fitted with a Xenon arc lamp for illumination. Two different cameras were tested, 1) the same Hamamatsu Orca Flash 4.0v2 used for the SIM system, and 2) a Blackfly BFS-U3-28S5M (FLIR). The pixel size of the Hamamatsu Orca Flash 4.0 is 6.5 μm whereas the pixel size for the Blackfly sensor was 4.5 μm. The same Nikon Plan Apo 10X 0.45NA objective lens was used in both systems.

2.2. Sample preparation and image acquisition

2.2.1. Fluorescent phantom beads

Green fluorescent spherical beads of about 10 μm diameter (Polysciences), with a coefficient of variation (CV) about 6%, and an excitation maximum of 488 nm were used in this work. The microspheres were diluted 100 times with phosphate-buffered saline, and well mixed. The sample (1 mL) was pipetted onto a glass slide, covered with a glass coverslip, and dried in a fume hood for 48 hours. We imaged the fluorescent beads with the 10x objective, with an exposure time of 25ms and a normalized spatial pattern frequency of 0.04ν. The intensity of the 470nm laser line was adjusted to prevent saturation of the CMOS camera. For beads imaging, we selected 20 regions on the slide and manually found their in-focus plane individually. We then moved the objective 50 μm away from the sample, and then systematically obtained z-stacks at each location with a step size of 5 μm towards the sample. In that case, for each location we had 21 layer z-stacks which spanned from out-of-focus to in-focus back to out-of-focus, with the central (11th) layer being consistently the most in-focus plane. Every location resulted in a 3D stack image with size 2048*2048*21 pixels. We cut the original image into 64 patches with size 256*256*21 to reduce the computation complexity. The raw data from each image capture at each z-layer included three patterned-illumination wide-field images, with the pattern shifted by one-third of the pattern period between each image.

2.2.2. Prostate samples

Fresh, intact human prostates were obtained right after radical prostatectomy surgery from patients providing informed consent under Tulane University Institutional Review Board approved protocols. The specimens were stained with 0.004% acridine orange (Sigma Aldrich) for 40 seconds and washed in phosphate buffered saline (PBS) by dipping for 15 seconds. Mosaic images of the prostate circumference were obtained with the structured illumination microscope. Approximately 3,000 4.2 megapixel 2D SIM images were obtained for each prostate in 50 minutes using 470nm laser illumination. In each case, we obtained four panels, and each panel (representing one circumferential aspect of the prostate) contained around 700 2D images each with a dimension of 2048*2048 pixels.

In addition, prostate core-needle biopsies from patients undergoing prostate diagnostic biopsy were also obtained from patients providing informed consent under Tulane University IRB approved protocols. Samples were stained with DRAQ5^{TM} far-red DNA fluorescent probe dye, (Biostatus, Ltd.) and Eosin Y (Leica Biosystems) and imaged within minutes of removal. For these samples, only images of the eosin fluorescence were used (using 532nm laser illumination).

2.3. Image acquisition

We reconstructed each single optically-sectioned SIM image from three shifted patterned images using the square-law detection algorithm described by Neil, et al. [10], then obtained perfectly co-registered wide-field (uniformly illuminated) images by averaging the three patterned images.

2.4. Network setup

In this work, Pix2Pix model was used from Isola, et al. [24] and implemented by Tensorflow 2.1, an open-source deep learning package. The whole pipeline is shown in Fig. 1. Data was augmented to prevent overfitting by applying a random jitter and mirroring the training dataset as a pre-process. The detailed parameters are shown in Fig. S2 and Fig. S3.

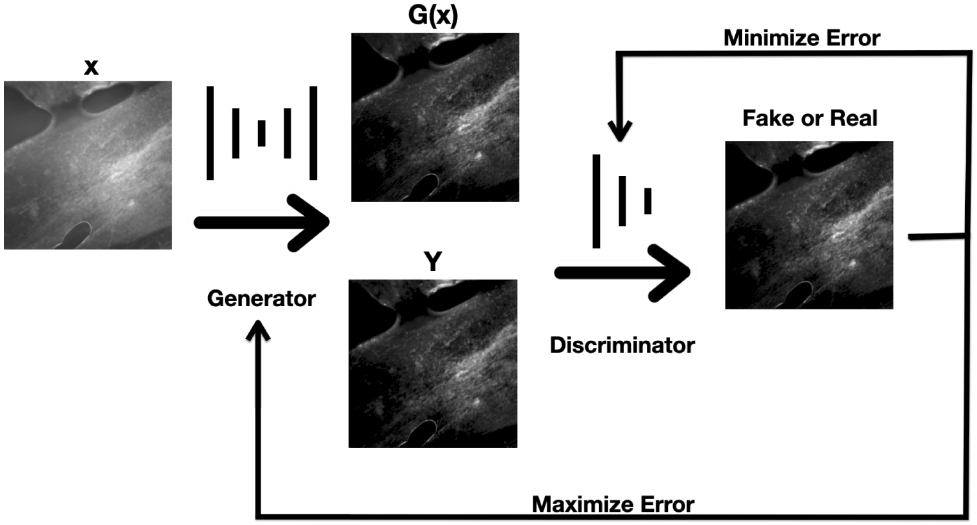

Fig. 1. Architecture of the model.

We paired each image using the computed SIM image X as the target/ground truth image and the corresponding wide-field image as the input image. The input image went through a generator G, which is an auto-encoder model based on U-Net architecture. The generator loss Eq. (1) is designed with a sigmoid cross-entropy loss of the generated images $G(X)$ and a matrix of ones, plus the mean absolute error MAE (L1 loss) between the generated images $G(X)$ and target SIM images Y multiplied by a lambda factor, with m referring to each pair of images.

$$G_{loss} = \frac{1}{m}\sum_{i=1}^{m} \log(1 - G(X)) + (\lambda_1 * MAE(G(X), Y)) \tag{1}$$

Next, the generated image $G(X)$ is paired with the corresponding ground truth image Y to go through the PathGAN [24] discriminator, mapping the image to an array with size N*N, trying to distinguish the forged N*N patch distribution from a real one. The discriminator is designed give a binary output to denote whether the patch is fake (0) or real (1). Additionally, the discriminator loss Eq. (2) is the sum of a sigmoid cross-entropy loss of the target images Y and a matrix of ones (real), plus a sigmoid cross-entropy loss of the generated images $G(X)$ and a matrix of zeros (fake).

$$D_{loss} = \frac{1}{m}\sum_{i=1}^{m}\log(1 - Y) + \log(G(X)) \tag{2}$$

To compare our results against traditional CNN, we used the CNN model from [17] using the same pre-processing methods, U-Net parameters, and learning rate as the cGAN model. In contrast to the CNN model, the generative network generates candidates while the discriminative network evaluates the candidates. The two neural networks tend to compete with each other to be updated dynamically, and the generator is not trained to minimize the distance to the ground truth image but to increase the error rate of the discriminator. The cGAN loss function attempts to minimize G_{loss} and maximize D_{loss} at the same time, Eq. (3). The overall Pix2Pix loss objective is shown in Eq. (4); here we set λ_2 to 100.

$$(G, D)_{loss} = E_{X,Y}[\log D(X, Y)] + E_X[\log(1 - D(X, G(X)))] \tag{3}$$

$$G*, D* = \arg\min_G \max_D (G, D)_{loss} + (\lambda_2 * MSE(G)) \tag{4}$$

The training was performed on a PC with Intel Xeon CPU E5-2620 @ 2.00GHz x 24 and 16GB RAM with an NVIDIA GeForce RTX 2080 Ti/PCle/SSE2 11GB GPU, assembled with CUDA version 10.1 on Ubuntu 19.10.

2.5. Checker-board effects

During parameter tuning, we noticed that there existed "checker-board" artifacts on the patches as shown in the left of Fig. 2, due to the nature of the deconvolution operation [39]. The uneven overlapping occurs when the kernel/filter size is not divisible by the stride, and gets multiplied along the length and width of the image. Moreover, the effect is much more severe in the areas containing little structural information, and increases exponentially over multiple layers due to the cascading. We tried several ways to solve this artifact: 1) we checked the log files of the training to make sure that the generator L1 loss and discriminator loss were not under 0.69, which would be seen as a not-converging failure for a GAN model generally [40]; 2) to train long enough such that the artifacts may disappear, we tried epochs up to 2,000, and saw that the artifacts were not alleviated but the loss began to rise; 3) we enlarged the receptive field of the discriminator, since if the discriminator looks at too myopic a region, it won't notice the textures are repeating [24]; 4) we replaced the transposed convolution in the up-sampling part of the U-Net with nearest neighbor plus traditional 2D convolution according to Distill et,al. [39], since the transposed

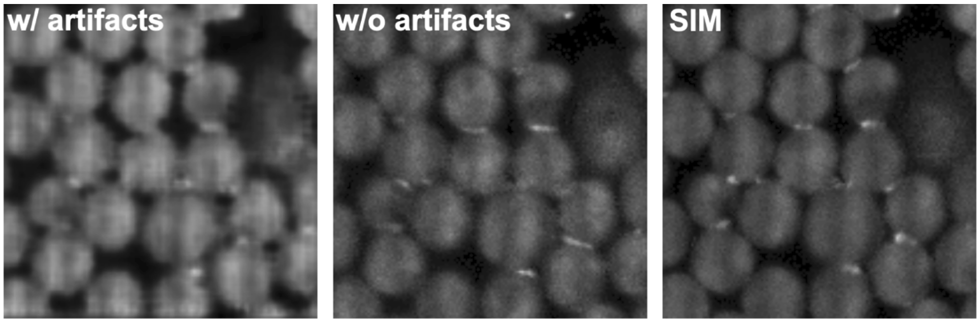

Fig. 2. Demonstration of "checker-board" artifacts (left) that are subsequently removed in our approach (middle). The resulting images are artifact-free and noise characteristcs are similar to ground-truth SIM images (right).

convolution will cause the uneven overlapping pattern; 5) we applied a Fourier transform to the predicted image, noticing the spectrum artifact which could be removed by a low pass filter; 6) we removed the batch normalization layer from the residual blocks, to normalize the mean and variance of each residual block layer, to clip the flexibility to the generator [23]; and 7) we improved the quality of both input and ground truth images. After several attempts, we combined method 3, 4 and 7 to modify the model to get a better result as shown in the middle panel of Fig. 2. As shown, this results in images free of these artifacts with similar noise characteristics to the ground truth images. This approach was therefore used for all GAN-predicted images presented in the results.

3. Results

3.1. Fluorescent phantom beads in 2D

Initially we demonstrated the cross modality capability of the Pix2Pix model by applying on phantom beads. We trained the 2D phantom beads model on 1,600 images with a patch size of 256*256 pixels. We randomly selected 90% for training, and tested on the remaining 10%. The model converged quickly in 300 epochs within three hours with a learning rate $2E-4$, and the training was terminated when both the generator loss and discriminator loss stabilized. The cGAN model was also compared to the traditional CNN model proposed by Zhang, et al. [17] by applying to the same training and testing dataset, with exactly the same pre-processing methods, U-Net parameters and learning rate. Two randomly selected patch examples of fluorescent microspheres (phantom beads) with wide-field (input), SIM (ground truth), CNN and GAN predicted images for representative 256*256 patches are shown in Fig. 3. The wide-field (WIDE) images contain blur from areas contaminated by out-of-focus light, especially within the thick spot where the beads have more than one layer (the bright spot). The optically-sectioned SIM images remove the out-of-focus light, making the in-focus layer more distinguishable. The CNN model seems to combine the wide-field and SIM images together, as the predicted phantom beads have a uniform shape with clear boundary, while the cGAN model tends to mimic the SIM images strictly, including SIM-reconstruction striping artifacts present in the ground-truth images. The yellow arrows in the first example indicates that the bead in the out-of-focus plane present in the wide-field image was removed by optical-sectioning both in the ground-truth SIM and predicted GAN images, but was erroneously preserved in the predicted CNN images. The yellow arrows in the second row show an example where SIM and GAN correctly preserved a bead in the focal plane, while the CNN obscured it. Also, the contrast is improved in the GAN vs. CNN predicted images, especially when the beads are overlapping.

Figure 4 shows the evaluation metrics for the CNN and cGAN models. The details of the evaluation metrics are provided in the Supplement 1. We calculated the A) MSE, B) SSIM, C) MI and D) PSNR for GAN vs. SIM, CNN vs. SIM, on over 100 randomly selected test patches and plotted the mean and standard deviation for each measurement. Generally, lower MSE and higher SSIM and MI indicate higher similarity between the two images. The cGAN model consistently demonstrated superior performance in evaluation metrics compared to the CNN model. We ran a two-sided T-test for the two approaches (GAN vs. SIM, CNN vs. SIM) in the four evaluation metrics, obtaining p-values of $1.27E-40$, $5.82E-50$, 0.101, and 0.003, respectively, illustrating there is a significant difference between the two approaches in MSE, SSIM, and PSNR, and approximately 10% chance that there is no statistical significant difference in MI. From these quantitative results we can conclude that the cGAN is superior or comparable to the CNN in the measures tested.

In order to validate that the predicted images would not add or remove important features that could affect the qualitative image interpretation, we chose phantom beads since we can count the actual beads number to quantify the prediction quality. We applied the standard marker-controlled watershed segmentation to detect the beads with following procedure [41] in MATLAB: 1)

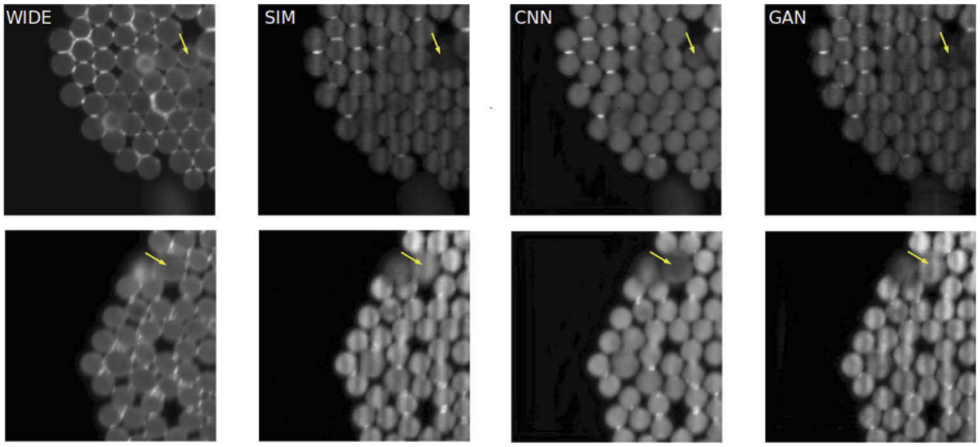

Fig. 3. Application of the 2D model to 10μm beads. Wide-field (WIDE), ground truth SIM, CNN-predicted and GAN-predicted images for representative 256*256 patches from two examples are shown from left to right. Yellow arrows indicate features that are more accurately reconstructed in the GAN images compared to the CNN images.

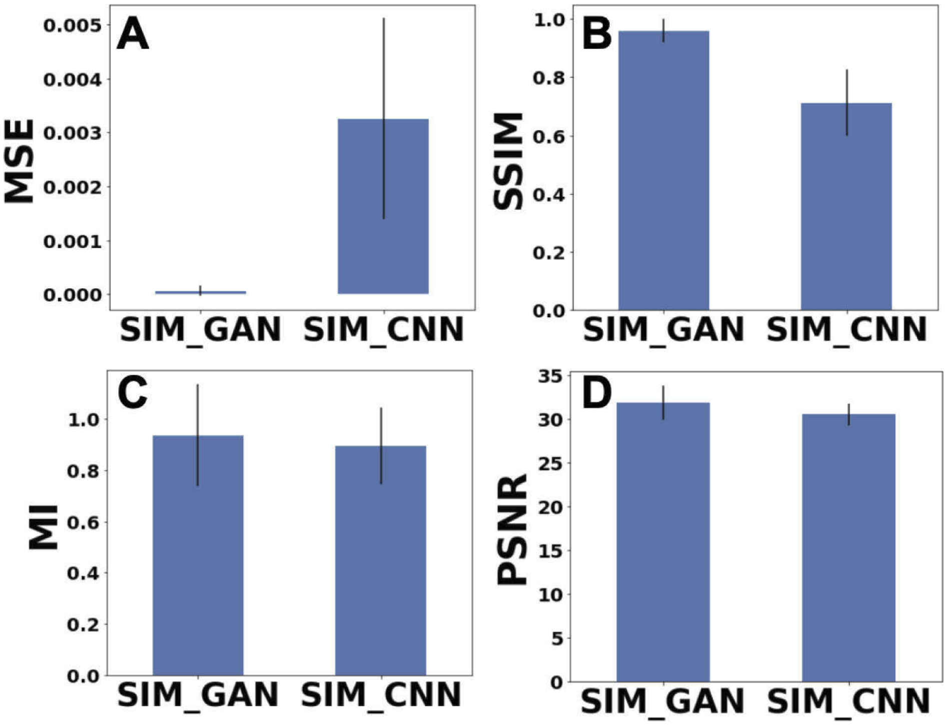

Fig. 4. Comparison of model accuracy compared to ground-truth SIM images for the models applied to phantom beads. A) MSE, B) SSIM, C) MI, D) PSNR. Modalities are as labeled: wide-field input image (WIDE), ground truth image (SIM), CNN predicted image (CNN), GAN predicted image (GAN).

compute the gradient magnitude to segment the objects; 2) mark the foreground objects by morphological opening and closing to find the regional maxima; 3) mark the background by

Otsu's thresholding; 4) modify the gradient magnitude to make sure the regional minima only occur at the foreground and background markers; 5) compute the watershed transform to return the segmentation; 6) measure the properties of each 8-connected component by regionprops function; 7) identify and count the round objects with a pre-defined diameter range. Figure 5 shows one random example for phantom bead counting by the watershed algorithm [42,43] for a SIM image (bead number=28) and corresponding GAN predicted image (bead number=29). Each blue dot denotes the detection of one phantom bead. From the example, the counting algorithm applied to the predicted image counted one more bead than the SIM image, located in the center of the beads cluster - this was likely because of the improved contrast between the beads and background in the GAN image. It should be noted that the actual bead count in both images would be the same at 29 if counted manually. The bottom left panel of Fig. 5 displays the bead counts calculated by the watershed algorithm for 200 test patches, with predicted bead counts plotted versus SIM ground truth bead counts. There is a high degree of correlation between bead counts in predicted and ground truth images. Since the watershed algorithm itself introduces non-avoidable errors, we also manually counted the bead number for 50 pairs of ground truth images and predicted images with the a priori knowledge of the bead size and bead shape, shown in the bottom right of Fig. 5. We strived to treat each test image with the same standard to avoid unintentional bias. The maximum difference in counts was three beads, and the two-tailed t-test P-value equals 0.2293, indicating no significant difference in bead counts between ground-truth and predicted images.

3.2. Prostate imaging

In order to have more realistic example to demonstrate the capability of the cross-modality image translation model from wide-field images to SIM images, we also tested it on human prostate tissue. We trained the 2D model on an image set of fresh, intact prostate surface tissue, using wide-field images as input and paired SIM images as ground truth for representative 2048*2048 pixel images. From the dataset we randomly selected 90% of the 3,000 images, and tested on the remaining 10%. We stopped the training as the model converged in 300 epochs in around 13 hours, with a learning rate $1E-4$.

The cGAN model was also compared to the traditional CNN model proposed by Zhang, et al. [17] by applying to the same training dataset. It took around 12.5 hours to train 400 epochs with the same learning rate $1E-4$. To evaluate the effectiveness of optical sectioning of the CNN and GAN models, two randomly selected patch examples from the test dataset are shown in Fig. 6. From the left to right column are shown the wide-field input image (blue outline), SIM image (red), CNN predicted image (yellow), and GAN predicted image (green). The first row shows two different patches, and the second row contains corresponding detailed images of the rectangles of the first row, where we emphasized a region of the patch to improve visibility. The bottom row shows the pixel intensity plot of the four different modalities along the marked lines from the two examples, and the residual of the pixel intensity plot between SIM and CNN (light blue), and SIM and GAN (fuchsia). From the result, we can see that the wide-field images are brighter since they represent all of the fluorescence emission, while SIM images suppress the emission above and below the focus plane and the contrast is improved. The CNN and GAN predicted images look similar to the SIM image, but the CNN tends to add more non-existing regular textures and noise features, particularly in the dark areas. We also reach the same conclusion from the line plots, where the pixel intensity of the wide-field images (blue) are high and have low-contrast peaks, while the others have higher contrast and lower overall baseline intensity. The absolute residual between GAN and SIM (light blue) is lower than for CNN and SIM (fuchsia). In general, the cGAN model produces the closest predicted image to the ground truth image, including more accurate background rejection and preserving SIM-reconstruction artifacts present in the original training pairs.

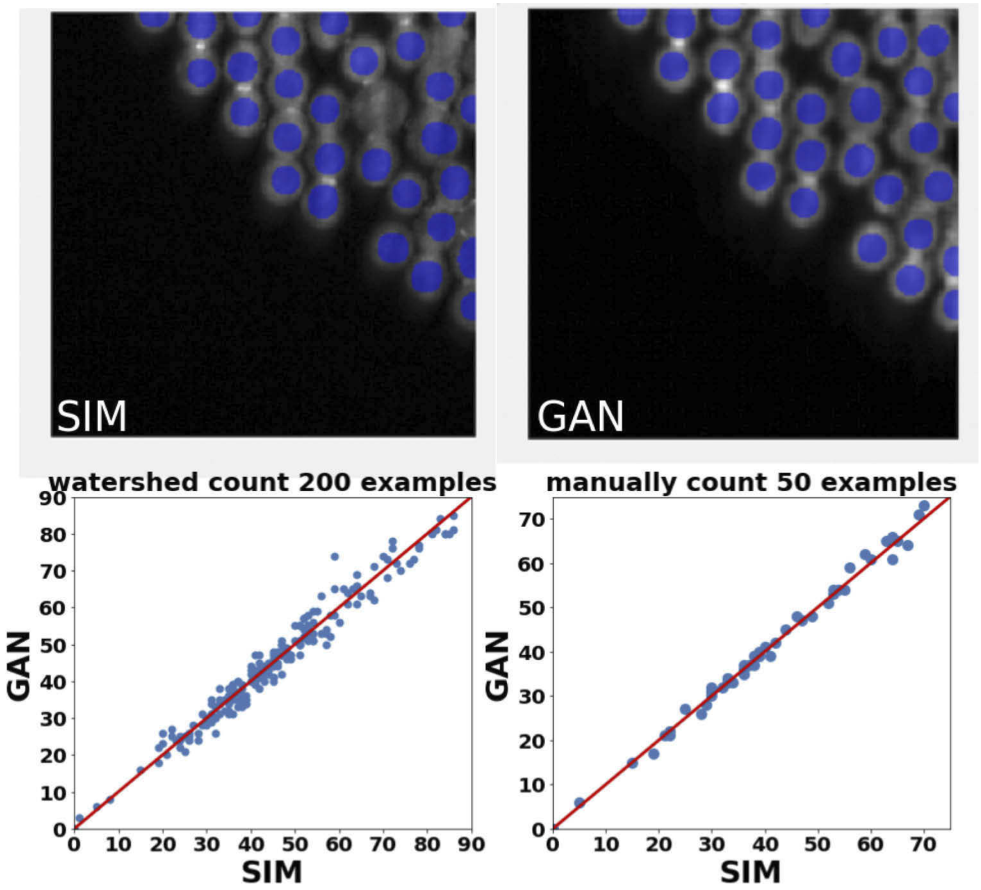

Fig. 5. Phantom beads counts. From left to right: mean phantom beads counts by watershed algorithm for ground truth (beads number=28), phantom beads counts by watershed algorithm for predict image (beads number=29), scatter plot for beads counts of each test patch with ground truth versus predict beads counts.

To quantitatively show the result, we plotted the mean and standard deviation of 50 randomly selected test patches in three evaluation measurements among the wide-field input image, SIM ground truth image, and two model predicted images in Fig. 7. The SSIM between SIM and GAN images are all above 0.85, while the SSIM between SIM and CNN images are below 0.8. Similarly, the MSE between SIM and GAN is lower, and the SSIM, MI and PSNR are higher. The result indicates that similarity of virtual SIM images predicted by the GAN model exceed those images predicted by traditional CNN model, as compared to the ground truth SIM images. Still, we ran the two-sided T-test across the two samples with P-values $8.37E-6$, $7.81E-3$, $1.05E-6$, $9.56E-3$ indicating there is statistical significant difference in those four evaluation metrics between GAN-SIM and CNN-SIM, in favor of higher accuracy for GAN.

Since the radical prostatectomy imaging application involves reconstruction of thousands of 4.2 megapixel images and stitching them into large mosaic images, we also developed a pipeline to take multiple adjacent images and to stitch them together into a larger mosaic to overcome computational or memory limitations, which is illustrated in the Supplement 1 (Fig. S4 and Fig. S5).

To validate the applicability of the model to novel image sets, we applied the well-trained model to a new dataset of wide field images from a different prostate specimen not used in

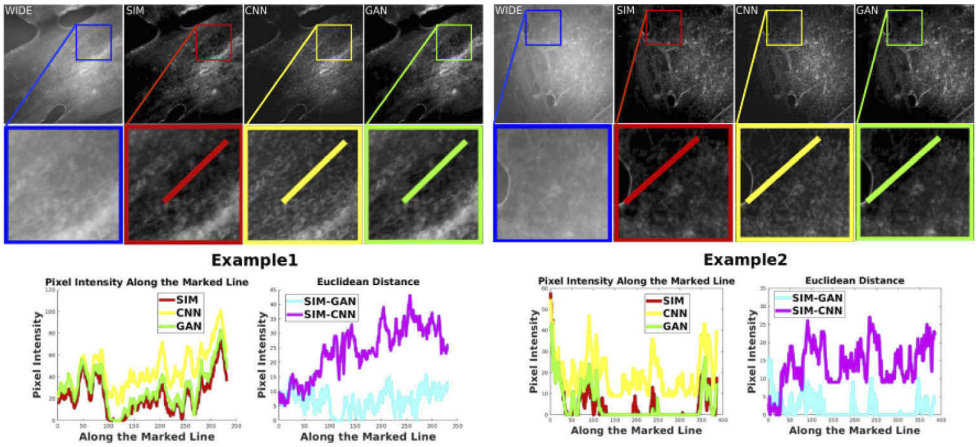

Fig. 6. Two examples of prostatectomy 2D imagery. From left to right: wide-field input image, SIM ground truth image, CNN predicted image, GAN predicted image. The bottom row shows the pixel intensity plots and residuals of pixel intensity plots along the marked lines for each example.

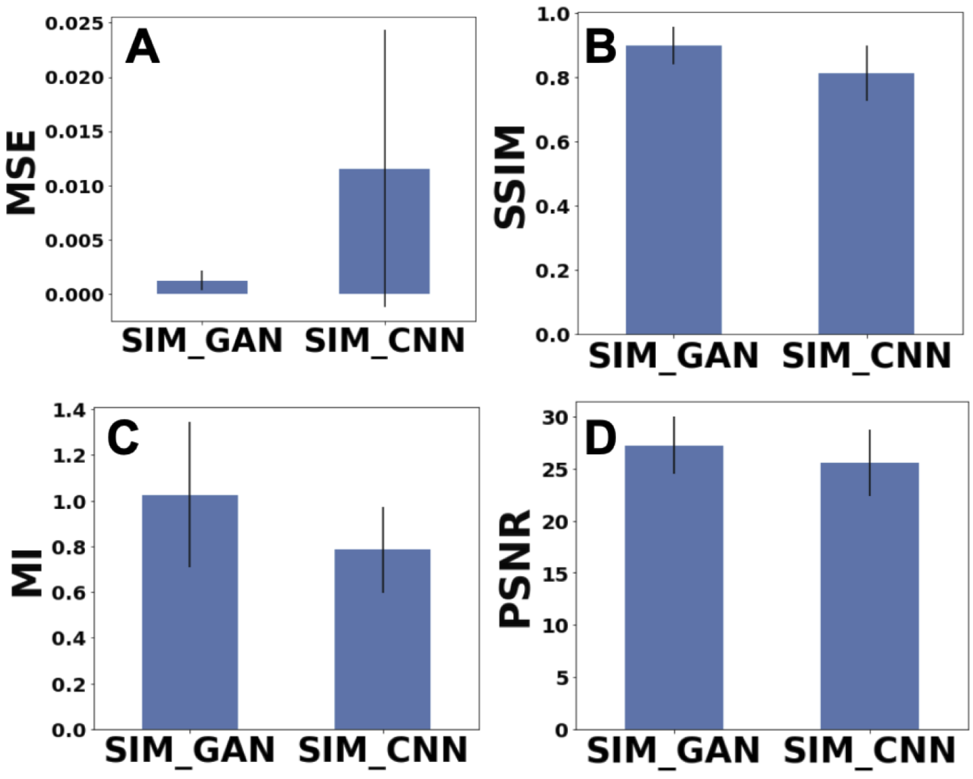

Fig. 7. Comparison of model accuracy compared to ground-truth SIM images for the models applied to radical prostatectomy images. A) MSE, B) SSIM, C) MI, D) PSNR. Modalities are as labeled: ground truth image (SIM), CNN predicted image (CNN), GAN predicted image (GAN).

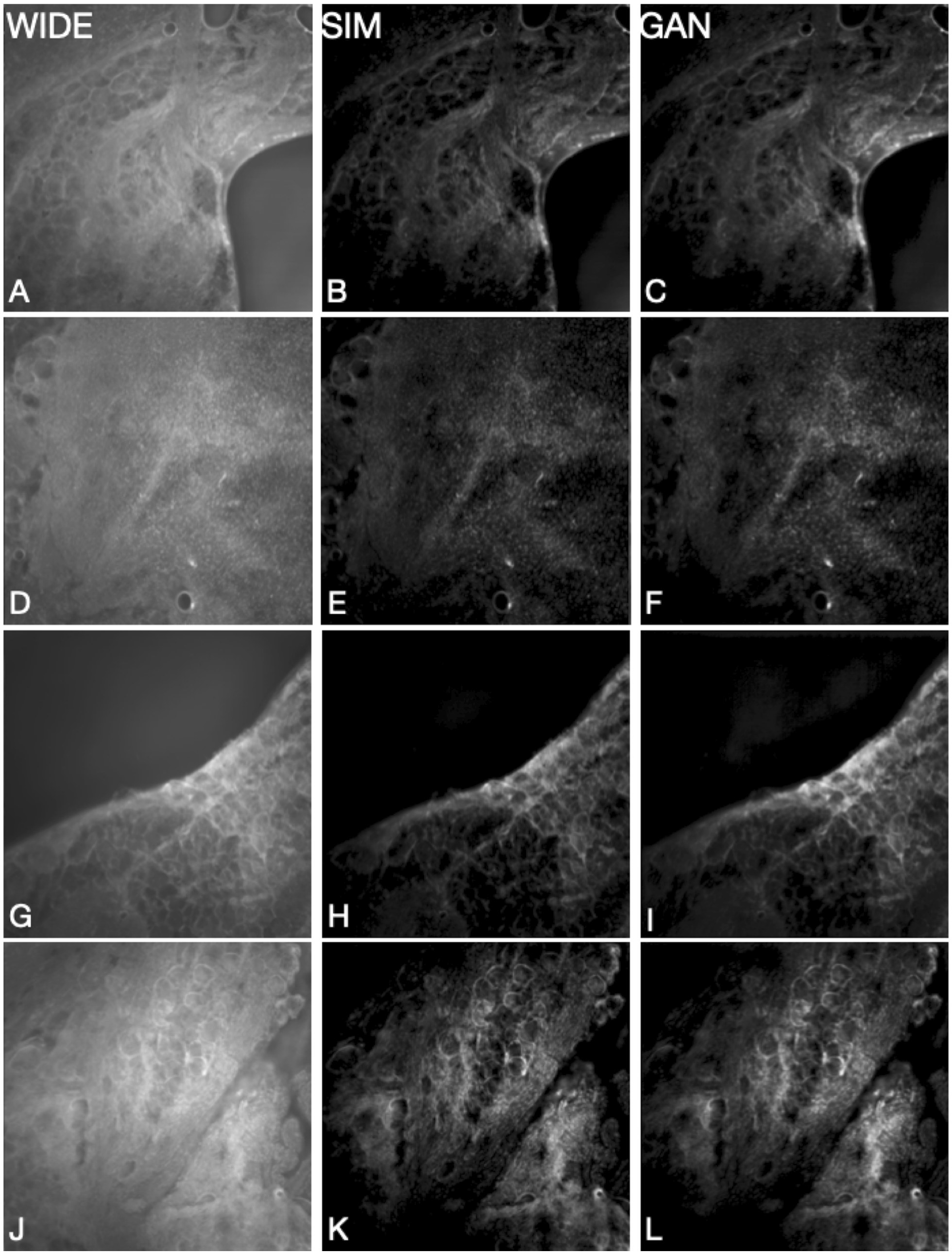

Fig. 8. Application of the cGAN model to radical prostatectomy (prostate surface) images from a patient not included in the training dataset. From left to right: wide-field input images (WIDE, A, D, G, J), SIM ground truth images (SIM, B, E, H, K), , and GAN predicted images (GAN, C, F, I, L).

the training set. The results are presented in Fig. 8, and demonstrates that the model can be applied to novel datasets, as long as the wide-field images are taken under the same condition and from the same sample type. Also, we tested the generalizability of the model to different sample types by applying the well-trained model to prostate biopsy (representative of prostatic parenchyma vs. prostate surface) tissues stained with a different fluorescent dye (shown in Fig. S6). Although the accuracy of the predicted tissue features was lower than for the same tissue type, we did demonstrate that the models appear to largely recreate the background rejection of optical sectioning by SIM; increased accuracy of tissue structure reconstruction could possibly be improved by including a wide range of tissue/sample types in the training datasets.

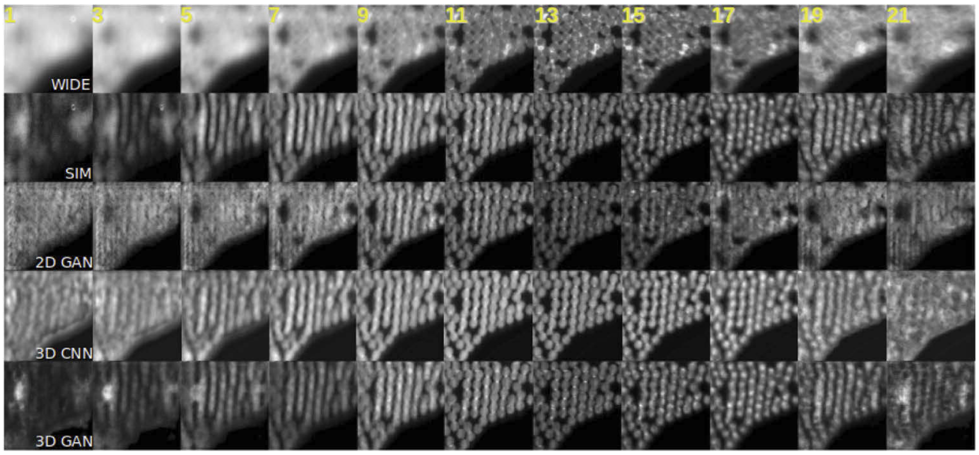

Fig. 9. Application of the 2D models and 3D model to 3D image stacks of the 10μm beads. One example of fluorescent microspheres, from top to bottom: wide-field (input), SIM (ground truth), predicted images trained on 2D cGAN model, predicted images trained on 3D CNN model and predicted images trained on 3D cGAN model of 21 odd layers from left to right) for representative 256*256 patches.

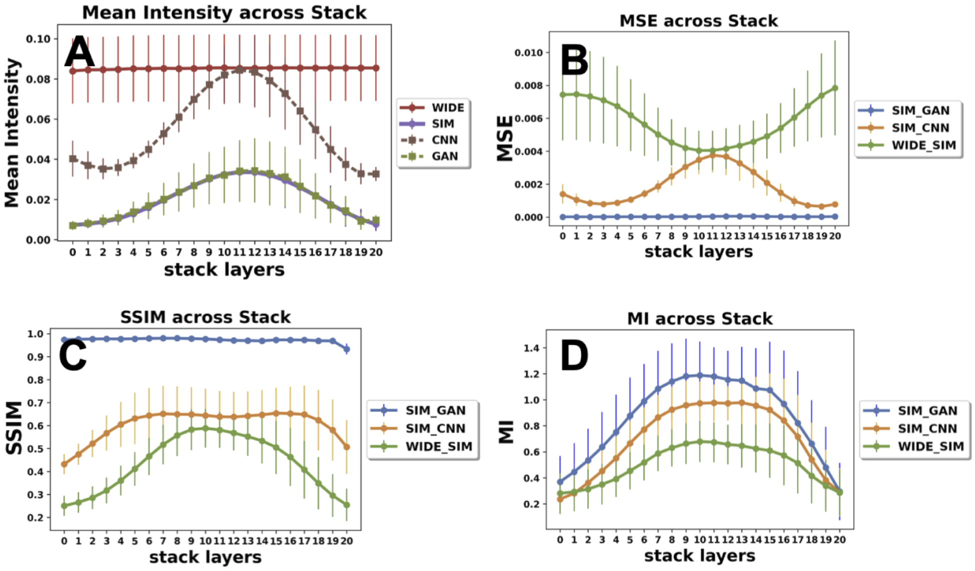

Fig. 10. Evaluation metrics for the 3D models. A) mean intensity of one patch across the 21 steps in the z-stack in four modalities; B) mean square error (MSE); C) structural similarity index (SSIM); and D) mutual information (MI) across the 21 steps in the z-stack between modalities.

3.3. Extension to 3D image reconstruction

We next tested whether the 2D model is capable of reconstructing true 3D imaging similarly to hardware-based optical sectioning modalities. We fed all 21 layers of a 3D stack of images of fluorescent beads into the 2D model from above, which was trained on only the center in-focus (11th) axial layer of the beads. Figure 9 shows one example (odd layers from left to right column) of wide-field (input) images, SIM (ground truth) images, and the predicted images for each of the layers in the z-stack. As we expected, SIM imaging removes the out of focus fluorescence emission in axial layers above and below the beads to achieve true 3-dimensional imaging. However, since the 2D model is trained only on the 11th layer, as expected only the predicted image of the 11th layer is similar to the corresponding SIM image, whereas the other axial layers are increasingly different from the corresponding SIM layers with increasing distance from the central in-focus layer. Specifically, the 2D cGAN predicted intensity is uniform across all z-stack layers, even as the predicted structure changes. This result indicates that we need a 3D model to replicate the behavior of true 3D optical sectioning.

Thus, we fed all 21 layers from the 3D stack images into the CNN and cGAN models for training, treating each layer as a different channel, with other parameters remaining the same. The 3D model converged within 500 epochs in around 20 hours. The fourth and fifth row of Fig. 9 shows the result of the CNN and cGAN 3D models compared to the 2D models and ground truth SIM images. The 3D cGAN model most faithfully preserves the axial imaging response shown in the ground truth SIM image stack, compared to the 3D CNN and 2D models.

We calculated evaluation metrics across the z-stack of of 30 randomly selected patches as shown in Fig. 10. Figure 10(A) shows the mean intensity of one patch across the 21 axial layers in the z-stack. As expected, the mean intensity remains constant through the stack for the input images (since there is no difference in intensity with axial defocus in wide-field imaging), whereas the ground truth SIM images manifested the behavior of 3D optical sectioning, in which the mean intensity is highest in the most in-focus layer, and decreases as the layers become out-of-focus both above and below the sample. The predicted 3D CNN and cGAN axial intensity curves revealed this optical sectioning behavior as well, although the SIM and GAN curves are virtually identical, in contrast to the CNN curve. Figure 10(B)-D also shows the mean square error (MSE), structural similarity index (SSIM) and mutual information (MI), respectively, across the 21 axial layers in the z-stack. As the MSE is closer to zero, SSIM closer to one, and the higher the MI, the perceptual difference between two images is less significant. As shown, the 3D GAN model demonstrated superior performance to the 3D CNN model in terms of both MSE and SSIM, with no discernible change in either metric with axial layer (confirming that the 3D GAN model faithfully recreates the optical sectioning behavior of SIM). In contrast, the 3D CNN model not only showed higher MSE and lower SSIM across all axial layers, but there was clear structure to the metrics across axial layers, showing enhanced performance at the in-focus layers, but poorer performance in axial layers away from the optimum focus. These results suggest that the CNN model does not accurately reproduce optical sectioning behavior, even when trained with axial stacks as for the more accurate 3D cGAN model.

3.4. Generalizability to traditional wide-field microscopes

Finally, we tested the generalizability of the model across microscope systems. Specifically, we used wide-field images from a Nikon TE2000 inverted fluorescence microscope that did not contain any of the hardware needed for SIM imaging. This microscope used a Xenon arc lamp for traditional Kohler wide-field illumination. The same microscope objective (Nikon Plan Apo 10X 0.45 NA) was used. However, we tested 2 different camera configurations: 1) we used the same Hamamatsu Orca Flash 4.0v2 camera used in the SIM system, and 2) we used a different FLIR Blackfly CMOS camera with smaller pixel size and different overall sensor size.

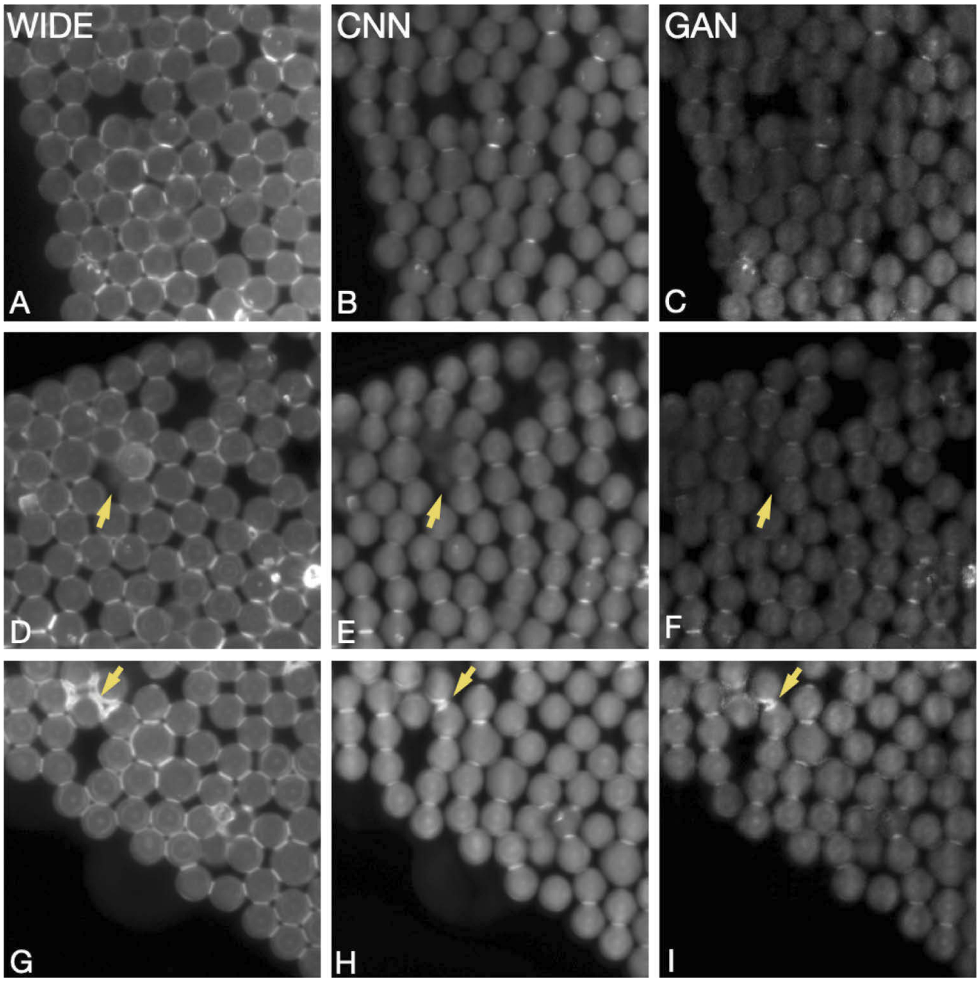

Fig. 11. Application of the model to wide-field phantom bead images collected using a different fluorescence microscope, but with the same camera. From left to right: wide-field input images (WIDE, A, D, G), CNN predicted images (CNN, B, E, H), and GAN predicted images (GAN, C, F, I).

We imaged phantom beads with the different microscope configurations and used the wide field images as inputs to both the cGAN and CNN models. Figure 11 shows the result with the same CMOS camera (Hamamatsu C11440-22CU, with 6.5 μm pixel size). The results show that both models are capable of optical sectioning. However, we did notice differences. The bottom row of the figure shows an example where both CNN and GAN achieved the task to remove overlapping emission from adjacent beads. However, in the middle row in this example, the cGAN model successfully removed a region of out-of-focus emission whereas the CNN preserved it. These behaviors are consistent with prior observations on the predictions from the SIM system, suggesting that the model may be generalized to other standard microscope systems using the same camera.

Moreover, Fig. 12 shows the result when a different camera is used (Blackfly BFS-U3-28S5M, with 4.5 μm pixel size). Similarly, as illustrated above, the first example (top row) shows that the GAN correctly predicts the intensity change with variations in focus across the image (left to right). However, in the CNN the predicted bead intensity is more uniformly distributed across

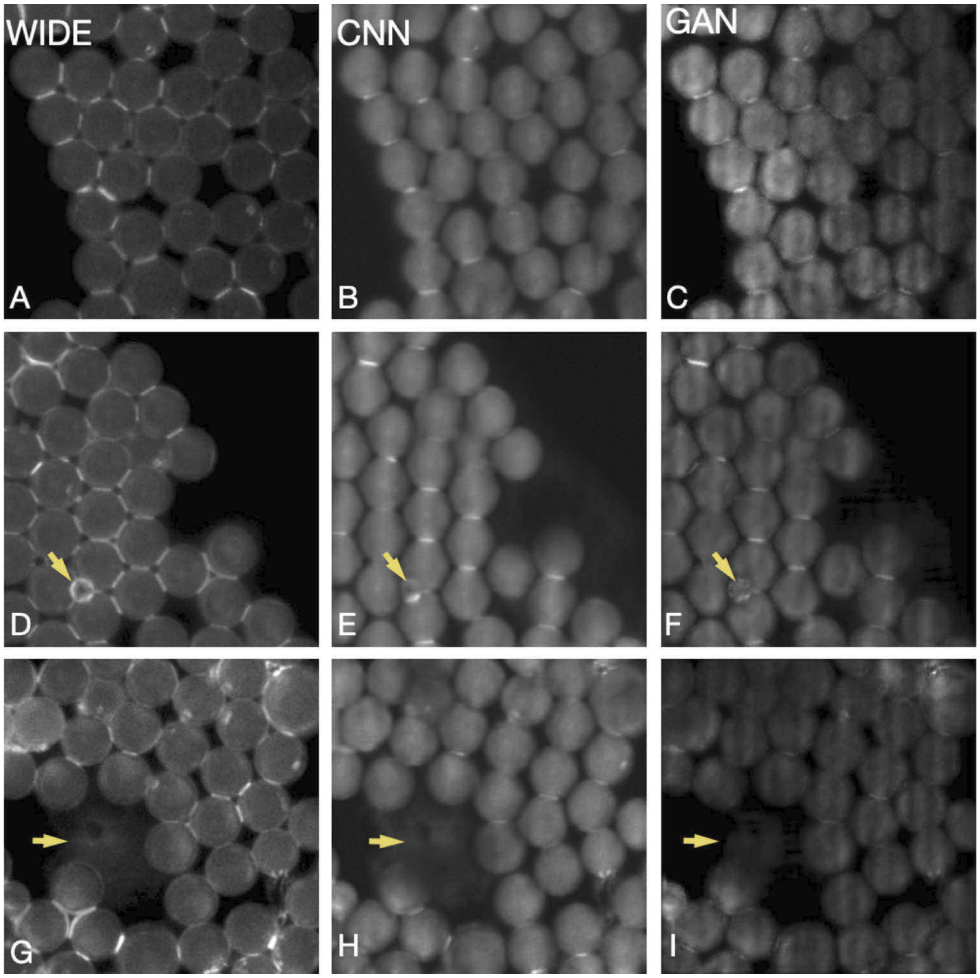

Fig. 12. Application of the model to wide-field phantom bead images collected using a different fluorescence microscope and with a different camera. From left to right: wide-field input images (WIDE, A, D, G), CNN predicted images (CNN, B, E, H), and GAN predicted image (GAN, C, F, I).

the image regardless of focus level. In the second example (middle row) the small bead resting between two larger beads is more accurately sectioned in the GAN predicted images vs. the CNN images. Finally, in the third example (bottom row) there is a region of noisy background emission in the wide-field image that is completely removed (as expected) in the GAN image, but is retained in the CNN image. Overall, these results suggest that the cGAN model is applicable to different microscope systems even with different cameras and pixel sizes while still maintaining a quality advantage over traditional CNN.

4. Discussion

Our work demonstrated that the Pix2Pix cGAN model, trained using paired wide-field and structured illumination microscopy images, can replicate optical sectioning using wide-field images as model inputs. We tested a 2D model, trained using in-focus wide-field and SIM image pairs, and applied it to both phantom image beads as well as fluorescently stained human prostate tissue. We found that the 2D cGAN model faithfully replicated the appearance of corresponding SIM images using only wide-field images as inputs, both qualitatively and quantitatively. To compare the use of Pix2Pix cGAN models to previous work using CNNs, we also employed a 2D CNN model to phantom bead and human prostate images. We demonstrated that the 2D cGAN model was superior to the 2D CNN model in terms of quantitative metrics. To alleviate bothersome checker board artifacts, we enlarged the receptive field of discriminator to avoid the repeating global texture, and eliminated the transposed convolution operations which cause the uneven pattern. We further demonstrated a pipeline for seamless reconstruction of large image mosaics, which we satisfactorily applied to large scale mosaics from human prostate tissue.

While the 2D models demonstrated excellent performance when in-focus wide-field images were used as inputs, in future practical applications of digital optical sectioning microscopy, it may be desirable to have prediction models that are robust to the defocus level of the input image. As expected, the 2D cGAN model did not perform well when input images were applied with greater defocus levels than existed in the training data. Therefore, we trained a 3D cGAN model using axial z-stacks from phantom bead data, and demonstrated that it faithfully recreated the behavior of true optical sectioning with axial defocus. Interestingly, we demonstrated that the 3D cGAN model greatly outperformed an alternate 3D CNN model both in terms of qualitative appearance and quantitative metrics.

There are a number of limitations to this work. At present, it is necessary to train the models with pairs of wide-field and optically sectioned images of the type of sample that one wishes to recreate with digital optical sectioning. Although obtaining perfectly co-registered pairs of wide-field and optically-sectioned images is a simple proposition in structured illumination microscopy, it does require that such data be collected for training for each type of sample of interest. It may be possible to obtain such training data using a structured illumination microscope, and then apply the prediction models on wide-field images from non-SIM systems with identical wide-field imaging specifications. In fact, our results indicate that this deployment strategy is feasible, as we demonstrated that the model could generalize to wide-field microscopes without any SIM optical components, including when a camera with different format size and pixel size was used. In addition, our results suggest that it is important for the most accurate reconstruction of sample structure to include 'like' samples in the training sets. Interestingly, although the model was less successful at predicting novel tissue structures not present in the training sets, it did appear to replicate the background rejection of optical sectioning, suggesting that generalizabilty to different sample types is possible with appropriate training image sets. Alternately, approaches such as Cycle-GAN [44] may be studied to determine whether it is possible to apply this method to a diverse set of datasets without pre-selection of datatype. These will be tested in ongoing and future work.

However, despite these limitations, our results demonstrate that digital optical sectioning using Pix2Pix cGANs is feasible and more accurate than similar CNN models, and potentially useful to enable users with standard wide-field microscopes to obtain optical sectioning performance in 2D and 3D with access to an appropriately trained model.

Funding. National Cancer Institute (R01 CA222831); National Science Foundation (IIS 2136744); National Institute of General Medical Sciences (U54 GM104940).

Acknowledgments. We gratefully acknowledge Madeline Behr, Ivan Bozic, and L. Spencer Krane, M.D. for their assistance in the collection of prostate samples and imaging data. We gratefully acknowledge financial support for this work provided by grants from the National Cancer Institute of the National Institutes of Health, R01 CA222831, and the National Science Foundation, IIS 2136744. Additional support was provided by grant U54 GM104940 from the National Institute of General Medical Sciences of the National Institutes of Health, which funds the Louisiana Clinical and Translational Science Center. This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation. JQB is a shareholder and employee of Instapath, Inc., which however did not financially or materially support this work. No other authors have disclosures to report.

Disclosures. JQB: Instapath, Inc. (I,E)

Data availability. The related code, checkpoint, partial data underlying the results presented in this paper are publicly available on Github [45]. All the other data may be obtained from the authors upon request, in accordance with institutional policies.

Supplemental document. See Supplement 1 for supporting content.

References

1. T. Wilson, *The Role of the Pinhole in Confocal Imaging Systems* (Springer US, 1990), pp. 113–126.
2. W. A. Carrington, K. E. Fogarty, L. Lifschitz, and F. S. Fay, *Three-dimensional Imaging on Confocal and Wide-field Microscopes* (Springer US, 1990), pp. 151–161.
3. T. R. Corle, C.-H. Chou, and G. S. Kino, "Depth response of confocal optical microscopes," Opt. Lett. **11**(12), 770–772 (1986).
4. W. Denk, J. Strickler, and W. Webb, "Two-photon laser scanning fluorescence microscopy," Science **248**(4951), 73–76 (1990).
5. C. J. R. Sheppard, "Multiphoton microscopy: a personal historical review, with some future predictions," J. Biomed. Opt. **25**(1), 014511 (2020).
6. M. Weber, M. Mickoleit, and J. Huisken, "Chapter 11 - light sheet microscopy," in *Quantitative Imaging in Cell Biology*, vol. 123 of *Methods in Cell Biology* J. C. Waters and T. Wittman, eds. (Academic Press, 2014).
7. J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. K. Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," Science **305**(5686), 1007–1009 (2004).
8. M. Saxena, G. Eluru, and S. S. Gorthi, "Structured illumination microscopy," Adv. Opt. Photonics **7**(2), 241 (2015).
9. R. Heintzmann and T. Huser, "Super-resolution structured illumination microscopy," Chem. Rev. **117**(23), 13890–13908 (2017).
10. M. A. A. Neil, R. Juškaitis, and T. Wilson, "Method of obtaining optical sectioning by using structured light in a conventional microscope," Opt. Lett. **22**(24), 1905–1907 (1997).
11. M. Wang, H. Z. Kimbrell, A. B. Sholl, D. B. Tulman, K. N. Elfer, T. C. Schlichenmeyer, B. R. Lee, M. Lacey, and J. Q. Brown, "High-resolution rapid diagnostic imaging of whole prostate biopsies using video-rate fluorescence structured illumination microscopy," Cancer Res. **75**(19), 4032–4041 (2015).
12. C. Ling, C. Zhang, M. Wang, F. Meng, L. Du, and X. Yuan, "Fast structured illumination microscopy via deep learning," Photonics Res. **8**(8), 1350–1359 (2020).
13. L. Jin, B. Liu, F. Zhao, S. Hahn, B. Dong, R. Song, T. Elston, Y. Xu, and K. M. Hahn, "Deep learning enables structured illumination microscopy with low light levels and enhanced speed," bioRxiv (2019).
14. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds. (Springer International Publishing, Cham, 2015).
15. X.-J. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," arXiv:1603.09056 (2016).
16. C. N. Christensen, E. N. Ward, M. Lu, P. Lio, and C. F. Kaminski, "ML-SIM: universal reconstruction of structured illumination microscopy images using transfer learning," Biomed. Opt. Express **12**(5), 2720–2733 (2021).
17. X. Zhang, Y. Chen, K. Ning, C. Zhou, Y. Han, H. Gong, and J. Yuan, "Deep learning optical-sectioning method," Opt. Express **26**(23), 30762–30772 (2018).
18. K. Ning, X. Zhang, X. Gao, T. Jiang, H. Wang, S. Chen, A. Li, and J. Yuan, "Deep-learning-based whole-brain imaging at single-neuron resolution," Biomed. Opt. Express **11**(7), 3567–3584 (2020).
19. R. N. Abirami, P. M. D. R. Vincent, K. Srinivasan, U. Tariq, and C.-Y. Chang, "Deep cnn and deep gan in computational visual perception-driven image analysis," Complexity **2021**, 1–30 (2021).
20. J. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman, "Toward multimodal image-to-image translation," CoRR arXiv abs/1711.11586 (2017).
21. M. Liu, T. M. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," CoRR arXiv abs/1703.00848 (2017).
22. X. Huang, M. Liu, S. J. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," CoRR arxiv abs/1804.04732 (2018).
23. Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for image-to-image translation," CoRR arXiv abs/1704.02510 (2017).
24. P. Isola, J. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," CoRR arXiv abs/1611.07004 (2016).
25. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," arXiv:1406.2661 (2014).
26. G. Nadarajan and S. Doyle, "Conditional generative adversarial networks for h&e to if domain transfer: experiments with breast and prostate cancer," Proc. SPIE **11603**, 116300 (2021).
27. E. A. Burlingame, A. A. Margolin, J. W. Gray, and Y. H. Chang, "SHIFT: speedy histopathological-to-immunofluorescent translation of whole slide images using conditional generative adversarial networks," Proc. SPIE **10581**, 1058105 (2018).

28. Z. Xu, C. F. Moro, B. Bozóky, and Q. Zhang, "GAN-based virtual re-staining: A promising solution for whole slide image analysis," CoRR arXiv abs/1901.04059 (2019).
29. G. Nadarajan and S. Doyle, "Realistic cross-domain microscopy via conditional generative adversarial networks: converting immunofluorescence to hematoxylin and eosin," Proc. SPIE **11320**, 113200S (2020).
30. P. Salehi and A. Chalechale, "Pix2pix-based stain-to-stain translation: a solution for robust stain normalization in histopathology images analysis," arXiv:2002.00647 (2020).
31. H. Zhang, C. Fang, X. Xie, Y. Yang, W. Mei, D. Jin, and P. Fei, "High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network," Biomed. Opt. Express **10**(3), 1044–1063 (2019).
32. H. Li, C. Zhang, H. Li, and N. Song, "White-light interference microscopy image super-resolution using generative adversarial networks," IEEE Access **8**, 27724–27733 (2020).
33. C. Fu, S. Lee, D. J. Ho, S. Han, P. Salama, K. W. Dunn, and E. J. Delp, "Fluorescence microscopy image segmentation using convolutional neural network with generative adversarial networks," CoRR arXiv abs/1801.07198 (2018).
34. M. Majurski, P. Manescu, S. Padi, N. Schaub, N. Hotaling, C. Simon, and P. Bajcsy, "Cell image segmentation using generative adversarial networks, transfer learning, and augmentations," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, (2019).
35. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nat. Methods **15**(12), 1090–1097 (2018).
36. F. Wang, T. R. Henninen, D. Keller, and R. Erni, "Noise2Atom: unsupervised denoising for scanning transmission electron microscopy images," Appl. Microsc. **50**(1), 23 (2020).
37. Y. Wu, Y. Rivenson, H. Wang, Y. Luo, E. Ben-David, L. Bentolila, C. Pritz, and A. Ozcan, "Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning," Nat. Methods **16**(12), 1323–1331 (2019).
38. Y. Wu, Y. Rivenson, H. Wang, Y. Luo, E. Ben-David, L. A. Bentolila, C. Pritz, and A. Ozcan, "Deep-z: 3D virtual refocusing of fluorescence images using deep learning," *Conference on Lasers and Electro-Optics* (2020).
39. A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," Distill http://doi.org/10.23915/distil (2016).
40. P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with a conditional GAN," arXiv:1611.07004 (2017).
41. K. Parvati, B. S. Prakasa Rao, and M. Mariya Das, "Image segmentation using gray-scale morphology and marker-controlled watershed transformation," Discrete Dynamics in Nature and Society **2008**, 384346 (2009).
42. A. Bieniek and A. Moga, "An efficient watershed algorithm based on connected components," Pattern Recognit. **33**(6), 907–916 (2000).
43. J. M. Sharif, M. F. Miswan, M. A. Ngadi, M. S. H. Salam, and M. M. bin Abdul Jamil, "Red blood cell segmentation using masking and watershed algorithm: a preliminary study," in *2012 International Conference on Biomedical Engineering (ICoBE)* (2012), pp. 258–262.
44. J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," CoRR arXiv abs/1703.10593 (2017).
45. Z. Huimin, B. Summa, J. Hamm, and J. Q. Brown, "Cross modality image translation from wide-field to SIM by pix2pix GAN: Code," Github 2021, https://github.com/zzzghm/wf2sim.git.